A Fast Mode Decision Algorithm for Downscaled Transcoding of H.264 Preencoded Video

Matthias von dem Knesebeck, *Student Member, IEEE*, Panos Nasiopoulos, *Member, IEEE*
University of British Columbia, Vancouver, Canada

***Abstract*-- In this paper, we present a fast mode decision method for transcoding pre-encoded H.264 video to a lower resolution. The algorithm reduces the computational load of the mode decision process during transcoding by adaptively considering information from the residual of the pre-encoded video and some intermediate results obtained in the re-encoding process. A performance improvement of 37% has been achieved compared to the best existing downscaling transcoding process known as the area-weighted vector median filter (AWVM).**

I. INTRODUCTION

With the continuous evolution of versatile mobile devices, the demand for mobile video applications is rapidly gaining popularity. At the same time, mobile devices suffer from innate constraints such as limited processing power, energy reservoir, storage space and transmission bandwidth. Due to these constraints, it is highly desirable to provide the content being displayed in a format that is optimized for the target device. However, existing content has often been pre-encoded with a resolution or bitrate exceeding these constraints, aiming at a different market, e.g., home entertainment.

Transcoding is a process that allows converting a bitstream from one format to another, in the present case from a high resolution to a lower resolution. The computationally expensive cascaded transcoding scheme involves decoding the original preencoded stream, downsizing the obtained frames using for instance a bilinear filter and subsequently re-encoding these downsized frames using an H.264 encoder. This process can consume a considerable amount of time and computational resources. Cascaded transcoding doesn't take advantage of existing information in the preencoded stream about compression characteristics such as motion vectors, modes or residual values. Hence, various methods have been proposed that re-use or consider that information in order to speed-up the re-encoding process.

Motion search and mode decision account for up to 90% of the computational efforts in video encoding.

For addressing the challenge of motion search, a number of algorithms have been proposed that allow re-estimating a suitable (initial) motion vector from the preencoded stream and hence reduce the burden of the element of motion search. Those include the simple average (SA), the activity-weighted average [1, 2], the area-weighted average (AWA) and the area-weighted vector median (AWMVM) [3]. The obtained motion vectors from these schemes are close approximations to the optimal motion vector. With them as a starting point, an additional 1-pixel motion search with a subsequent quarter-pixel refinement is still required to obtain high compression efficiency. AWMVM gives the best results among the above methods and is used as a base for our proposed method.

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the British Columbia Innovation Council.

H.264 defines seven block-sizes for every temporally predicted (INTER-) frame (16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4) as well as a SKIP mode, which does not encode a residual signal. When considering a downscaling ratio of 2:1, four 16x16 blocks of the original preencoded stream will be reduced to one 16x16 block in the transcoded stream. Each block can itself consist of a number of sub-partitions.

In order to identify the most efficient choice among these block sizes, H.264 aims to minimize a Lagrangian cost function for every coding mode:

$$J = D + \lambda R \tag{1}$$

where D denotes the distortion between the predicted and the original macroblock, R represents the number of bits used to encode the motion information, and λ is the Lagrangian multiplier that is itself dependent on the quantization factor chosen. The coding mode resulting in the lowest cost J is finally selected for encoding.

In order to reduce the computational burden of mode decision that is due to evaluating all possible coding modes, a number of methods have been proposed for encoding from raw video and have been applied for transcoding [3-6]. However, improved performance may be achieved when additional information from the preencoded video stream is taken into account, as it is subsequently presented in our algorithm.

II. PROPOSED ALGORITHM

The goal is to find a measure that can be obtained with little computational effort and that would allow an early decision about the optimal mode for the current macroblock.

After transcoding a number of test sequences with the cascaded scheme, we analyzed the distribution of the block sizes within every frame of the resulting downsized video.

We have found that there is a correlation present between the distribution of large, medium sized and small partition sizes within every 16x16 macroblock and the sum-of-absolute residual values (*SAR*) found within the areas in the preencoded stream that correspond to these macroblocks (in case of a 2:1 downsizing ratio, those corresponding areas contain 32x32 pixels). The correlation indicates that small partition sizes are found in the downscaled stream where the absolute residual values in the preencoded video are large and vice versa. However, this statement does not hold true in all cases.

When re-encoding the downscaled video using the AWMVM as a starting point for the 1-pixel motion search, we also found that a second measure, the distribution of the cost values J obtained from the 16x16 mode search within a frame exhibits a significant correlation as well with the distribution of the resulting final block sizes chosen.

Interestingly, our analysis unveiled that in the vast majority of cases when the first measure failed to indicate the best block size, the second measure would make a correct prediction and vice versa. We concluded that a combination of these two measures might yield the desired predictor that fulfills the two goal criteria. The following linear combination of the two measures, obtained by running an optimization algorithm, minimizes the error between the predicted block size distribution and the desired block size distribution and yields an overall very strong correlation:

$$M_i = 1.25 \cdot SAR_i + J_i, \quad i \in \{1..N\} \tag{2}$$

with M_i denoting the measure that is used to assign a set of valid modes for the given (16x16) macroblock i within a frame with N macroblocks, SAR_i indicating the sum-of-absolute residual values in the corresponding area of the preencoded video and J_i stating the cost value obtained from the 16x16 mode search.

With this measure at hand, we now need to define the decision criteria to assign the appropriate set of modes to the current macroblock. While the distribution of M remains strongly correlated with the block-size distribution for varying motion content, the mean and the range of the values M_i within a frame change with the content. We concluded that a criterion involving the mean (μ) and the standard deviation (σ) would define a suitable factor for this task. A close analysis unveiled that the adaptive thresholds in Table 1 are most suitable for assigning a set of valid modes to every macroblock i.

TABLE 1: VALID MODES FOR EVERY MACROBLOCK

Threshold for M_i	**Valid Modes**
$M_i > \mu+0.5\sigma$	all modes
$\mu+0.5\sigma > M_i > \mu-\sigma$	16x16,16x8,8x16
$M_i < \mu-\sigma$	16x16 only

III. EXPERIMENTAL RESULTS

The proposed algorithm was implemented using the JM reference software 14.2 on a 3.0 GHz Pentium IV platform. Simulations were performed with the first 100 frames of three standard test sequences (Akiyo, Foreman, Mobile) using IPPP GOP structure and 1 reference frame.

Table 2 shows the results for transcoding the three test sequences from CIF to QCIF resolution with the AWMVM method and the proposed algorithm, detailing PSNR, bitrate, encoding time, number of search points visited per MB and the number of pixel comparisons performed per MB (CMP/MB).

We observe that 37.68% of pixel comparisons are avoided by using the proposed method with a small decrease in picture quality by 0.04dB and a bit rate increase of 0.92% on average.

Fig. 1 illustrates the rate-distortion performance of the proposed algorithm and the AWMVM method for the obtained QCIF resolution stream of the Foreman sequence.

TABLE 2: EXPERIMENTAL RESULTS (CIF -> QCIF), QP 20

Sequence	**Method**	**PSNR (dB)**	**bitrate (kbps)**	**Enc.Time (s)**	**SrchPts per MB**	**CMP/MB**
Foreman	AWMVM	42.44	346.4	808.1	5,148	150,172
	Proposed	42.40	350.4	636.5	2,070	92,677
	Δ	-0.04	1.17%	-21.23%	-59.79%	-38.29%
Akiyo	AWMVM	44.49	77.4	577.2	4,395	125,418
	Proposed	44.45	77.6	448.3	1,974	81,175
	Δ	-0.04	0.30%	-22.34%	-55.09%	-35.28%
Mobile	AWMVM	41.40	1,109.8	647.2	5,209	140,641
	Proposed	41.35	1,124.1	500.1	2,184	85,124
	Δ	-0.05	1.28%	-22.74%	-58.07%	-39.47%
Average		**-0.04**	**0.92%**	**-22.10%**	**-57.65%**	**-37.68%**

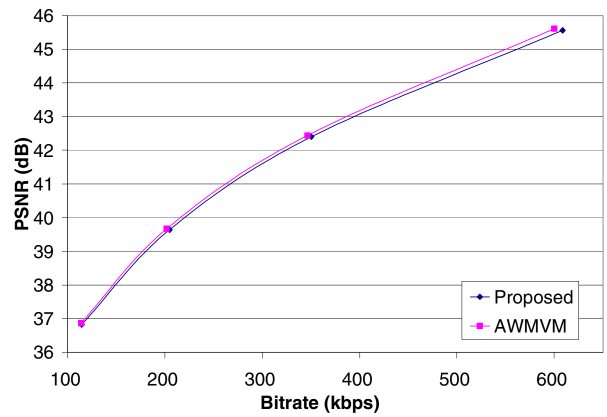

Fig. 1: Foreman (CIF → QCIF)

IV. CONCLUSION

We presented an early-termination mode decision process for efficiently transcoding H.264 video content to a smaller resolution by a factor of 0.5. The method takes advantage of correlation present between the residual of the preencoded stream, the final cost value from the 1-pixel full search around the AWMVM vector and the set of optimal coding modes for a given macroblock. Our performance evaluations have shown that about 37% of the computations can be saved compared to the unmodified AWMVM method while preserving picture quality (-0.04dB) and incurring a 0.92% increase in bitrate.

REFERENCES

[1] B. Shen, I.K. Sethi and B. Vasudev, "Adaptive motion-vector resampling for compressed video downscaling", *IEEE Trans. on Circuits and Syst. for Video Technology*, vol. 9, pp. 929-36, 1999.

[2] M. Chen, M. Chu and S. Lo, "Motion vector composition algorithm for spatial scalability in compressed video", *IEEE Trans. on Consumer Electronics*, vol. 47, pp. 319-25, 2001.

[3] Y. Tan and H. Sun, "Fast motion re-estimation for arbitrary downsizing video transcoding using H.264/AVC standard", *IEEE Trans. on Consumer Electronics*, vol. 50, pp. 887-94, 2004.

[4] C. Liu, W. Wei, N. Yang and C. Kuo, "Motion vector re-estimation for video trans-coding with arbitrary downsizing," in *Int. Conf. on Intellig. Inform. Hiding and Multimedia Signal Processing*, pp. 806-9, 2008.

[5] H. Shen, X. Sun, F. Wu, H. Li and S. Li, "A fast downsizing video transcoder for H.264/AVC with RD optimal mode decision," in *IEEE Int. Conf. on Multimedia and Expo*, pp. 2017-20, 2006.

[6] P. Zhang, Y. Lu, Q. Huang and W. Gao, "Mode mapping method for H.264/AVC spatial downscaling transcoding," in *IEEE Int. Conf. on Image Processing*, pp. 2781-4, 2004.